Part I

Introduction and Theory

Chapter 1

Introduction

1.1 Motivation

In recent years, understanding and quantifying the global hydrologic cycle has become a priority research topic. Soil moisture, in particular, has gained a lot of attention as it constitutes a key variable in the global hydrologic cycle. Land-atmosphere processes critically depend on the state of soil moisture, as soil moisture partitions the energy fluxes available at the land surface into latent and sensible heat fluxes. In addition, soil moisture conditions are important in determining the amount of infiltration and groundwater recharge.

Improving our understanding of soil moisture and temperature conditions will help us in many ways. Global circulation models are now routinely used in weather and climate predictions, but they usually contain only inadequate representations of the physical processes at the land-atmosphere interface. A better understanding of soil moisture and temperature dynamics will therefore help us with the assessment and prediction of global change and improve our ability to produce reliable short-term weather forecasts.

Sustainable management of water resources for agricultural and urban use will be feasible if we are able to more accurately quantify soil moisture conditions and the corresponding recharge into groundwater aquifers. Predicting floods is not only a question of knowing how much precipitation will reach the ground. An accurate flood forecast also depends on a good knowledge of the prevalent soil moisture conditions. Moreover, there is a feedback of soil moisture onto precipitation [Eltahir, 1998]. But to usefully incorporate such feedback mechanism into hydrologic and meteorologic predictions, including the forecast of droughts, it is again necessary to know the prevalent state of soil moisture.

Traditionally, improving models of large-scale soil moisture dynamics has been difficult due to the lack of corresponding large-scale observations. However, the advent of remote sensing data has now made it possible to study land-atmosphere processes on large spatial scales. Ideally, the satellite data are used in conjunction with the existing land-surface models to extract the valuable information contained in both the data and the models. Such optimal merging of data and models is generally termed data assimilation. In a variational assimilation scheme, the estimates are determined by minimizing a measure of fit between the land-surface states and both prior information and new data. The measure of fit is formulated using weights that depend on the corresponding uncertainties.

The observations that are available for assimilation are not always direct measurements of the land surface variables of interest. This is especially true for satellite remote sensing data. Satellites, for instance, cannot observe soil moisture directly, and only satellite-observed radiances may be used to infer soil moisture conditions. In this case, the model consists not only of a component for soil moisture dynamics, but also of a forward Radiative Transfer scheme relating the soil moisture fields to the remotely sensed radiances (or brightness temperatures). Alternatively, off-line soil moisture retrievals could be obtained by inverting the Radiative Transfer model. It is preferable, however, to assimilate satellite radiances directly. It is much easier to specify appropriate weights for radiances in the objective function than to estimate covariance structures for soil moisture retrievals. Moreover, the off-line inversion of the Radiative Transfer model presents an unnecessary source of error.

Unlike sparse and infrequent observations, which can only be related to particular fields at particular times and locations, estimates produced by an assimilation scheme can provide a complete description in time and in space of the entire land-surface state, including soil moisture, soil temperature, and canopy temperature. This is achieved by using a dynamic model as part of the data assimilation algorithm. The data are effectively interpolated in time and extrapolated in space by respecting the dynamical and physical constraints. From such a complete picture, land-surface processes can be examined in detail. In meteorology, the number of investigators using such estimated "data sets" is probably much greater than those using any individual data type [Errico, 1999].

It is important to emphasize the fact that data assimilation reaches beyond mere model calibration. An optimal data assimilation algorithm will consider all the useful information and the errors contained in the model and the data along with the corresponding error statistics. In addition to the consideration of measurement error, a modern assimilation scheme usually involves the assumption of imperfect models, which is reflected in parameter and model error (or process noise) terms in the model equations. Moreover, posterior error covariances can be inferred.

Whereas model calibration is typically implemented to estimate a set of parameters once and for all, data assimilation algorithms are designed to run in an operational mode, continuously estimating state variables of interest. An additional feature of modern data assimilation algorithms is the possibility to test scientific hypotheses by formulating the model together with the statistical assumptions for the errors as a null hypothesis. If the hypothesis is rejected, the data are not statistically consistent with the underlying assumptions on the model and the errors. In this case, the estimates are of little meaning, but we would have learned something about land-surface dynamics. Finally, data assimilation provides a valuable tool for assessing and validating observation systems.

Data assimilation techniques have been successfully used meteorology and oceanography. In meteorology in particular, data assimilation has led to considerable improvements in the quality of short-term weather forecasts over the past few decades. Today, six hour global forecasts of wind and temperature produced with estimates derived from data assimilation algorithms are generally as accurate in a root-mean-square sense as most individual verifying observations themselves [Errico, 1999].

Hydrologists now face the challenge to apply true data assimilation techniques to all problems where remote sensing data can provide new insights. However, this is a difficult task due to the highly nonlinear nature of land-surface processes, the size of the problem, and the lack of data and experience to determine error statistics accurately. Consequently, the implementation of data assimilation techniques always requires trade-offs between resolution, complexity, computational effort, and data availability.

This study is predominantly a feasibility study. Its main goals are (1) to develop an optimal land surface data assimilation algorithm, and (2) to determine how useful remotely-sensed L-band ($1.4GHz$) passive microwave measurements could be for the large-scale estimation of soil moisture.

1.2 Environmental Data Assimilation

In this Section, we briefly review a few influential studies in the field of environmental data assimilation with a focus on large-scale applications. Our primary goal is to introduce data assimilation techniques that have been used in the environmental sciences and to establish a broader frame of reference for this thesis. We certainly do not claim to provide a full review of the topic.

This Section covers a wide range of applications, mostly drawn from meteorology and oceanography. In meteorology, in particular, massive amounts of observations have been available for decades as operational data streams. Early on it has been indispensable to develop methods that make optimal use of these data for numerical weather forecasting and model development. More recently, large-scale operational observations that are useful for oceanographers have become widely available, and advanced data assimilation systems have been successfully developed and applied.

Large-scale hydrologic data assimilation, however, is still a field very much in its infancy. This probably owes as much to the scarcity of large-scale data as to the lack of consensus about how best to model land surface processes. In Section 1.2.3 we briefly present a few studies on hydrologic parameter estimation and data assimilation. A specific survey of soil moisture data assimilation can be found in Section 1.3.

Please note that our partitioning of the discussion into meteorologic, oceanographic, and hydrologic data assimilation does not at all imply that the methods used in these fields are different or separate. In fact, almost all assimilation techniques currently used in environmental data assimilation are simplifications or variants of the weak-constraint variational technique (Section 2.1) or, equivalently, the Kalman smoother [Gelb, 1974]. For details on the equivalence and approximations of the optimal methods consult the review papers cited below. All techniques can theoretically be applied to almost any dynamic problem in the geosciences, and the most important factor in determining the choice of method is usually computational feasibility.

1.2.1 Data Assimilation in Meteorology

"One of the main reasons we cannot tell what the weather will be tomorrow is that we do not know accurately enough what the weather is today. [. . .] Data at the initial time of a numerical forecast can be supplemented, however, by observations of the atmosphere over a time interval preceding it. New observing systems [. . .] make it absolutely necessary to find new and more satisfactory methods of assimilating meteorological observations — for the dual purpose of defining atmospheric states and of issuing forecasts from states thus defined". This quote is taken from the preface of a volume on progress in data assimilation published in 1981 [Bengtsson et al., 1981]. Almost two decades later, the European Centre for Medium-Range Weather Forecasts (ECMWF) has implemented a fully four-dimensional data assimilation algorithm in their operational forecast system [Klinker et al., 1999].

Since the early days of numerical weather forecasting, researchers have been trying to merge data and models. Excellent descriptions, reviews, and comparisons of the various data assimilation techniques used in meteorology have been provided by Le Dimet and Talagrand [1986], by Ghil and Malanotte-Rizzoli [1991], and by Daley [1991] on general methods, by Lorenc [1986] on variational methods, and by Todling and Cohn [1994] on sequential methods. Courtier et al. [1993] compiled a literature list on the use of adjoints, variational methods and the Kalman filter in meteorology. The list includes comments and goes back to 1955.

Weak-constraint Variational Assimilation

Unfortunately, truly optimal operational data assimilation with a full Kalman filter, a Kalman smoother or an equivalent variational technique[1] is still not computationally feasible, and the algorithm implemented at the European Centre (ECMWF) is not yet ideal. (Why this is so will be discussed below.) However, a successful large-scale research application of a fully optimal data assimilation approach has been presented by Bennett et al. [1996]. Their study is unique in that the model is only imposed as a weak constraint. In other words, errors in the model formulation are taken into account as process noise (or model error). The optimal estimate is derived with the variational representer approach, which is presented in detail in Section 2.3.

Bennett et al. [1996] invert a global Numerical Weather Prediction (NWP) model using about 2500 scalar data from reprocessed cloud-track wind observations. However, in an operational setting the assimilated data should include all of the global quality-controlled but otherwise raw observations. The authors point out that there are about 40,000 in situ observations alone [Daley, 1991], which clearly shows the current limitations of the technique in an operational context.

Strong-constraint Variational Assimilation

There have been numerous attempts at simplifying either the model equations (although the physics did not change, of course) or the optimal estimation equations in order to make operational data assimilation computationally feasible. The algorithm recently implemented operationally at the European Centre (ECMWF) is based on the variational scheme 4DVAR [Thépaut and Courtier, 1991; Thépaut et al., 1993; Courtier et al., 1994]. In 4DVAR, the model is assumed perfect and imposed as a strong constraint, that is model errors are neglected. Only uncertainties in the initial and boundary fields are taken into account. If model error is present, as is certainly the case, using the model as a strong constraint may result in erroneous adjustments of the estimates. In other words, the estimates of the initial and boundary conditions must compensate for any significant model errors. However, unlike the weak-constraint method proposed by Bennett et al. [1996], 4DVAR is already feasible in the operational environment of the European Centre (ECMWF).

Simplified Kalman Filters

4DVAR is certainly an improvement over conventional Optimal Interpolation [Rabier et al., 1993], which is used in most other weather forecasting centers (see below). But neglecting model errors does constitute a serious limitation. As an alternative to 4DVAR, one could

[1] Optimality refers to the implementation of the full Kalman filter, the Kalman smoother, or a weak-constraint variational algorithm. We ignore for a moment any suboptimality resulting from nonlinearities in the physics.

think of sophisticated approximations to the Kalman filter which allow for model errors but rely on a simplified propagation equation for the forecast error covariance. Such approximations are also called low-rank approximations of the full Kalman filter. Dee [1991] suggests a Kalman filter in which a simplified version of the dynamic model is used for the forecast error covariance propagation. Alternatively, Cohn and Todling [1996] suggest a filter based on partial eigenvalue decompositions of the forecast error covariance together with adaptive tuning based on reduced resolution ideas.

Optimal Interpolation

To this day, almost all operational weather forecast centers still use Optimal Interpolation or variants thereof as their method of choice. Optimal Interpolation can be viewed as a simplified Kalman filter in which the propagation of the estimation error covariance is entirely neglected [Daley, 1991]. In return for the computational savings, the complicated and time-dependent error covariance fields of the atmospheric states must be accurately estimated. In practice, this is quite impossible and leads to rather suboptimal assimilation algorithms, even though the name of the method would suggest otherwise. Moreover, Optimal Interpolation is usually implemented in the spectral domain, which limits the choices of error covariance models in practical applications. Finally, the approximate solution method for the update equations in conventional Optimal Interpolation can lead to dynamic imbalances [Cohn et al., 1998].

For these reasons, NASA's Data Assimilation Office has recently developed the Physical-space Statistical Analysis System (PSAS) as an improved variant of Optimal Interpolation [Cohn et al., 1998; Chen et al., 1999]. The new method has been included into the Goddard Earth Observing System (GEOS) data assimilation package. PSAS operates in physical space rather than in the spectral domain and employs a different numerical method to solve for the updates. Since it is operating in physical space, PSAS is capable of using more advanced error covariance models than conventional Optimal Interpolation.

1.2.2 Data Assimilation in Oceanography

In oceanography, operational data streams have not been available in the same way as in meteorology. Therefore the focus of the investigations has been somewhat different, oriented more towards learning about ocean dynamics by using optimal methods for individual case studies whenever data are available. Ghil [1989], Bennett [1992], Ghil and Malanotte-Rizzoli [1991], Evensen [1994a], Malanotte-Rizzoli [1996], and Wunsch [1996] offer good collections, descriptions, reviews, and comparisons of the various attempts to solve inverse problems in oceanography.

Weak-constraint Variational Assimilation

Egbert et al. [1994] use the direct representer algorithm (Section 2.3) to estimate global tides from the TOPEX/POSEIDON altimeter data. Even though the tide model is linear, which is rarely the case for geophysical applications, the number of remote sensing data is still too big for a naive implementation of the direct representer approach. Egbert et al. [1994] therefore develop a set of steps in which they reduce the dimensionality of the problem.

Eknes and Evensen [1997] extend the representer formalism by solving a simultaneous parameter and state estimation problem with a weak-constraint formulation for an Ekman model. Finally, Bennett et al. [1998] apply the indirect representer algorithm to assimilate data from the Tropical Atmosphere-Ocean (TAO) array into a coupled model of the tropical Pacific. They compare their weak-constraint approach with a strong-constraint algorithm and reach the important conclusion that the assumption of a perfect model must be clearly rejected in this case study.

Simplified Kalman Filters

A low-rank approximation of the Kalman filter has been applied by Verlaan and Heemink [1997] to the tidal flow forecasting problem. Their approach is to combine a reduced rank approximation of the error covariance with a square root factorization. The use of the factorization ensures that the error covariance matrix stays positive-definite at all times, while the smaller rank reduces the computational effort.

Asif and Moura [1999] develop a computationally efficient formulation of the optimal Kalman filter which is based on the block structure that results from the discretization of the partial differential equations commonly used in the physical sciences and from the sparseness of the measurements, for example satellite scans. The authors further develop an approximate implementation of the block Kalman filter. Underlying this simplified implementation is the approximation of the inverse error covariance matrix, that is the information matrix, by a sparse block banded matrix. Such banded approximations correspond in essence to modeling the error field in the spatial estimates at each point in time as a reduced-order Markov random field. To demonstrate the concept, Asif and Moura [1999] use the optimal filter and the simplified scheme to assimilate synthetic satellite altimeter data into a linear shallow water model. The comparison shows that the suboptimal filter performs well and that the approximations of the simplified filter are reasonable.

Yet another simplification to the Kalman filter for large-scale applications has been proposed by Evensen [1994b]. In the so-called Ensemble Kalman Filter, the error covariance is propagated with a Monte Carlo method. Instead of solving the Riccati equation for the error covariance evolution, the scheme is based on propagating an ensemble of model forecasts. If a measurement becomes available, the forecast error covariance needed for the update step is estimated from this ensemble. Evensen and van Leeuwen [1996] use the Ensemble Kalman Filter to assimilate Geosat altimeter data into a two-layer quasi-geostrophic model of the Agulhas Current. The validity of the ensemble approach is obviously dependent on the size of the ensemble. It does seem daring to estimate the forecast error covariance from an ensemble of 500 model trajectories when the state vector is approximately of dimension 100, that is when the error covariance matrix contains on the order of 10,000 elements.

Very recently, Lermusiaux and Robinson [1999a] presented an assimilation scheme based on a combination of the Ensemble Kalman Filter and a reduced rank approximation. Like in the Ensemble Kalman Filter, the error covariance is propagated with a Monte Carlo approach. Before the update step, however, the covariance matrix is reduced in rank. The authors formulate an objective criterion to decide whether the addition of another member to the existing ensemble is necessary or not. The result is a suboptimal filter which tracks an evolving error subspace in space and in time. Consequently, the scheme is termed error subspace statistical estimation (ESSE).

In a companion paper, Lermusiaux and Robinson [1999b] apply their algorithm to shelf-break front simulations in the Middle Atlantic Bight. Identical twin (synthetic) experiments are conducted under the assumption of a perfect model, that is model errors are neglected.

Moreover, the synthetic observations do not contain measurement error, although a small measurement error is used in the estimation algorithm. The proposed filter compares favorably to a traditional Optimal Interpolation scheme which does not include any error covariance propagation. In case the computational demand for the proposed filter is too big for a given operational application, the authors suggest that their scheme could be used to improve the parameterizations of the Optimal Interpolation approach.

Multiresolution Optimal Interpolation

Fieguth et al. [1995] have developed and applied a new variant of Optimal Interpolation. The goal is to provide interpolated estimates at multiple resolutions at the time of the update or analysis step. The multiresolution algorithm is a generalization of time series state-space models for which the Kalman filter is an efficient estimator. When applying the multi-scale estimation technique, the biggest task is to build a model for the particular application that fits the covariance matrix at the finest scale. In addition to providing interpolated estimates and accompanying error variance statistics at multiple resolutions, a striking advantage of the multi-scale estimation framework is that its complexity scales linearly with the problem size. Moreover, the efficiency of the algorithm is entirely insensitive to irregularities in the sampling or spatial distribution of measurements and to heterogeneities in measurement errors or model parameters. Consequently, the approach has the potential of being an effective tool in a variety of remote sensing problems.

Fieguth et al. [1995] have applied the multiresolution estimation algorithm to the interpolation and statistical analysis of the TOPEX/POSEIDON altimeter data in the North Pacific Ocean. Another application of the multi-scale Optimal Interpolation algorithm to the mapping of temperature in the northeastern Pacific has been published by Menemenlis et al. [1997]. The authors also concern themselves with the development of a class of multi-scale statistical models appropriate for oceanographic mapping. Finally, Fieguth et al. [1998] have applied the method to map the sea level anomaly of the Mediterranean Sea based on TOPEX/POSEIDON and ERS-1 data. Unfortunately, the extension of the multiresolution framework to problems with temporal evolution presents formidable challenges. The development of temporally dynamic models is the subject of ongoing research.

1.2.3 Data Assimilation in Hydrology

In hydrology, inverse methods have traditionally been focusing on parameter estimation and model calibration rather than state estimation. In particular for groundwater inverse problems, measurements are scarce, and highly heterogeneous parameters such as the hydraulic conductivity are virtually unknown a priori. McLaughlin and Townley [1996] offer an excellent review of the subsurface data assimilation problem. Also, Zimmerman et al. [1998] compare seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. Recently, Reid [1996] and Sun [1997] have worked on parameter estimation in groundwater contaminant transport problems. Finally, Daniel et al. [1999] have applied the multiscale estimation approach described in Section 1.2.2 to the estimation of solute travel time.

Hydrologic data assimilation as a state estimation problem has only very recently become a topic of widespread interest. In a review of hydrologic data assimilation published in 1995, McLaughlin [1995] is *"unaware of any studies which use distributed watershed models* *to assimilate field data".* Recent data assimilation efforts to estimate soil moisture are summarized in Section 1.3.

1.3 State of the Art of Soil Moisture Data Assimilation

In this Section, we attempt to assess the state of the art of data assimilation techniques used for the estimation of soil moisture. Only selected works will be discussed, with a focus on studies that apply optimal estimation techniques or at least non-trivial approximations thereof. By no means do we claim for our overview to be complete.

Two obvious classification schemes can be applied. In a first scheme, the studies can crudely be classified into a first category consisting of spatially one-dimensional physical models using synthetic data [Entekhabi et al., 1994; Milly, 1986], or small-scale field data [Mahfouf, 1991; Katul et al., 1993; Parlange et al., 1993; Galantowicz et al., 1999; Calvet et al., 1998; Callies et al., 1998; Bouyssel et al., 1999; Castelli et al., 1999], and a second category, in which large-scale field cases have been investigated [Houser et al., 1998; Bouttier et al., 1993b; Rhodin et al., 1999]. In the first category of small-scale studies, the dimensions of the state vector and the observation vector are small, and the computational effort for truly optimal estimation is easily bearable. The models in the second category of large-scale applications are horizontally distributed and of high dimensionality. Consequently, only suboptimal filters have been implemented to date.

A second classification could be based on the data types that are assimilated. With the exception of [Castelli et al., 1999], the studies of Sections 1.3.1 and 1.3.2 use either direct measurements of soil moisture or remotely sensed brightness data which are very closely related to surface soil moisture. Since such measurements of soil moisture are not yet available operationally, there have been numerous investigations on soil moisture data assimilation from low-level atmospheric parameters such as air temperature and relative humidity at $2m$ above the ground. However, these parameters are only weakly and indirectly related to surface soil moisture. The latter studies are geared towards improving numerical weather prediction and treat soil moisture rather as a tuning parameter. For this reason we describe them in the separate Section 1.3.3.

1.3.1 One-dimensional Optimal Estimation Approaches

If the modeled land surface system is one-dimensional and contains only a single vertical column, the dimension of the state vector is small and the application of truly optimal estimation techniques is not limited by computational resources. One such optimal technique is the Kalman filter [Gelb, 1974], which has been used by many investigators. Other investigators have applied a variational approach, which is described in detail in Section 2.1. Among the latter are Mahfouf [1991], Callies et al. [1998] and Bouyssel et al. [1999]. Since they assimilate low-level atmospheric observations to infer soil moisture, their studies are discussed in Section 1.3.3, even though optimal variational assimilation methods are used.

The Study by Milly [1986]

Milly [1986] presented a study to determine the optimal temporal characteristics of a remote soil moisture sensor. He uses a very simple linear soil moisture model in which the parameters are perfectly known. The forcing consists of a sequence of equally spaced Dirac delta functions to model the precipitation input. Milly [1986] uses a full Kalman filter to evaluate the relative merits of the accuracy and the sampling frequency of the measurements.

The Studies by Katul et al. [1993] and by Parlange et al. [1993]

Katul et al. [1993] use an Extended Kalman Filter (EKF) for the estimation of the soil moisture state in a simple bucket model. The model is obtained from the depth-integration of a one-dimensional version of Richards' equation, which results in a nonlinear state-space formulation with a scalar state. A no-flow boundary condition is imposed at the top, and the flux at the lower boundary is prescribed. The hydraulic conductivity and the soil moisture content are related through a simple exponential-type two-parameter model.

The assimilated soil moisture data are neutron-probe measurements from a small field drainage experiment carried out by the authors. In addition to the state estimation, Katul et al. [1993] also estimate the two soil hydraulic parameters, the initial estimation error variance, and the model error of the state-space formulation. These four parameters are determined through repeated runs of the Extended Kalman Filter. For every run, a set of parameters is guessed, and a goodness-of-fit objective function is evaluated. The goodness-of-fit is measured with a sum of squared differences between the predicted states and the corresponding measurements. No prior information about the parameters is used. The final set of parameters is then given by the best fit. To carry out the optimization, the authors implemented a simplex scheme.

In a similar study, Parlange et al. [1993] estimate the field scale diffusivity together with the initial estimation error variance and the model error of the state-space formulation. The starting point here is an approximate solution to the depth-integrated diffusion equation, combined with a water balance equation. Once the model equation is cast into a state-space formulation, the mechanics of the estimation algorithm are identical to the approach by Katul et al. [1993].

The Studies by Entekhabi et al. [1994] and by Galantowicz et al. [1999]

The studies by Entekhabi et al. [1994] and by Galantowicz et al. [1999] stand out because an optimal data assimilation approach is applied to a multi-layer model of soil moisture and temperature dynamics. The authors use a Kalman filter to update the temperature and moisture profile from observations of the brightness temperature. The spatially one-dimensional model is entirely physically-based, making use of Richards' equation, the heat equation, and a model for the radiative transfer.

Entekhabi et al. [1994] show that it is possible to infer information about the temperature and the moisture at depths below the penetration depth of the microwaves. Note, however, that the focus is on the methodology. Most importantly, the data are completely synthetic and vegetation is not modeled. Only one vertical column is considered. In addition, updates from the brightness temperature and the infrared temperature data are made hourly, which is not a very realistic situation.

In a very recent study, Galantowicz et al. [1999] present an assimilation algorithm which is based on the Kalman filter and similar to the one in Entekhabi et al. [1994]. The algorithm is tested on field data, namely data from the Beltsville Agricultural Research Center (BARC), Maryland, taken during a seven-day drydown in July 1994 [Jackson et al., 1997]. Moreover, the authors test their algorithm with a four-month series of simulated operational conditions. The results indicate that the soil moisture profile can indeed be retrieved from updates of the brightness temperature made only every three days, and that the proposed data assimilation scheme is stable.

The Study by Calvet et al. [1998]

Calvet et al. [1998] present a comprehensive study on the feasibility of retrieving root zone soil moisture from surface soil moisture or surface soil temperature observations. They use the ISBA (Interaction between Soil, Biosphere, and Atmosphere) surface scheme of the French weather forecast system, which models soil moisture in just two layers, a very shallow surface layer and a deep reservoir.

The assimilation technique is a strong-constraint variational method. The uncertain parameter is the initial soil moisture of the deep reservoir, and the objective function to be minimized consists of the root mean square difference between the measured and the simulated values of the observed surface soil moisture content. No prior regularizing term is included in the objective. The data are from two months of field observations taken in Spring and Fall 1995 in southern France. The assimilation period is either a moving fifteen-day window or a moving five-day window during the thirty-day observation periods. In a series of assimilation experiments, observations are available to the estimation algorithm from twice daily to once every four days.

In conclusion, Calvet et al. [1998] suggest that deep soil moisture can indeed be retrieved with reasonable accuracy from surface soil moisture observations once every three days, but concede that soil moisture estimation from soil temperature measurements can at best work under dry conditions. Finally, the authors conclude that the length of the assimilation window should not be less than ten days.

The Study by Castelli et al. [1999]

A major goal of the study by Castelli et al. [1999] is to reduce the data needs for surface flux and soil moisture estimation. Therefore, the authors only assimilate observations of ground temperature, which are readily obtained from current remote sensing platforms. The uncertain input is a time-dependent parameter which is called soil moisture index. The soil moisture index describes the limitation of evaporation due to the limited availability of soil water and is closely related to the surface heat flux.

Castelli et al. [1999] use a variational technique and include the surface energy balance as a physical constraint in the objective function. In mathematical terms, the estimation of the time-dependent soil moisture index amounts to the estimation of a state-dependent model error term. The scalar weights used in the objective function imply that this model error is not correlated in time.

Estimates of the surface heat flux and the soil moisture index are derived from the data of the First International Satellite Land Surface Climatology Project Field Experiment (FIFE). The experiments cover the summer months of 1987 and 1988, but the individual assimilation windows are limited to one day. Daily averages of the estimated surface heat flux compare well to independent latent heat flux observations. However, the authors conclude that there is a need to discriminate between soil moisture and aerodynamic contributions to the surface control over evaporation.

Discussion

The straightforward application of optimal data assimilation algorithms to one-dimensional problems has met with fair success, and the potential for inferring soil moisture from remote sensing observations of passive microwave data has clearly been demonstrated. In light of these results, the most pressing question is how best to extend the techniques presented above to large-scale applications.

1.3.2 The Study by Houser et al. [1998]

The so far most comprehensive study on soil moisture data assimilation has been carried out by Houser et al. [1998]. The authors modified and extended the TOPLATS land-atmosphere transfer scheme [Famiglietti and Wood, 1994a; Famiglietti and Wood, 1994b]. TOPLATS is a spatially distributed hydrologic model to predict the diurnal dynamics of the water and energy fluxes at the land surface as well as the local vertical recharge into the underlying aquifer. Its algorithms are intentionally simpler than the ones used in operational surface-vegetation-atmosphere transfer schemes (SVATS). The basic components of TOPLATS are water balance equations for the canopy and the soil as well as an energy balance equation at the surface. The original model describes the unsaturated zone with two layers, a root zone and a transmission zone. Houser et al. [1998] added a shallow third soil layer at the top. The soil moisture in this new soil layer can possibly be inferred from remote sensing. The soil hydraulic properties are parameterized with the model of Brooks and Corey [1964] and the soil moisture dynamics are based on an approximate analytical solution of Richards' equation using infiltration and exfiltration capacities [Eagleson, 1978]. Horizontal flow exists only in the underlying saturated layer. In the unsaturated zone, lateral flow is completely neglected. The model is applied to the Walnut Gulch watershed in southeastern Arizona.

In the following, we briefly describe the data assimilation techniques that have been applied by Houser et al. [1998]. In all cases, the assimilated data are soil moisture values that have been obtained through an off-line inversion of remotely sensed microwave observations.

Direct Insertion and Statistical Corrections

The simplest data assimilation method used by Houser et al. [1998] is Direct Insertion. Here, all observations are assumed perfect. In the update step, the model prediction is simply replaced with the measurement for all observed components of the state vector. No other assimilation is performed, nor are the observations pre-interpolated. This results in very abrupt discontinuities of the soil moisture field. Any advection of information is accomplished through the subsequent prediction steps. The propagation of the error covariance is entirely neglected. The computational savings are enormous, and the computational effort is almost the same as for a pure simulation run without using the data at all. However, the scheme is wholly suboptimal.

Houser et al. [1998] also employ a technique they call Statistical Corrections. In this approach, the mean and the variance of the observations are computed. Then the components of the predicted state vector are rescaled in order to match the statistics of the observations.

Nudging

Next, Houser et al. [1998] test two forms of nudging. The idea behind nudging is to add an artificial forcing term to the model equation such as to drive the state continuously towards the observations. This artificial forcing is not obtained from covariance propagation and the filter is thus suboptimal. Two nudging techniques are used. In "Nudging Towards a Gridded Analysis", the observations are pre-interpolated to the model grid. This means that the scheme can only be applied *within* a region of observations. After the pre-interpolation, "observations" are available for every state. In the second nudging technique, termed "Nudging to Individual Observations," no interpolation is carried out. In both nudging techniques, the artificial forcing is entirely empirical. Houser et al. [1998] implemented all possible combinations of the two nudging techniques together with the method of Statistical Corrections. The authors were most satisfied with "Nudging to Individual Observations" both inside and outside the region of observations.

Optimal Interpolation

Optimal Interpolation (or Statistical Interpolation) is a special case of the Kalman Filter [Daley, 1991]. In Optimal Interpolation, the error covariance propagation equations are omitted. Houser et al. [1998] approximate the predicted error covariance (or background error covariance) for the computation of the gain in the following way. First, the difference between observed and TOPLATS-simulated values is computed. Second, the covariance of this difference is calculated. Third, an analytical covariance model is fitted to the data. Houser et al. [1998] use a model for a climatological background derived by Thiebaux [1976]. Although Optimal Interpolation is always less computationally demanding than a full Kalman Filter, the computational effort for the application of Houser et al. [1998] was still by far too large. The authors therefore reduced the number of measurements in an ad hoc fashion. They followed two approaches. In the first approach, a random subset of observation is chosen for each grid point, and together with the observation closest to the given grid point, only these measurements are assimilated. All other measurements are discarded. In the second approach, "super-observations" are obtained by averaging the available measurements over a coarser spatial grid and thus reducing their number.

Discussion

Except for Optimal Interpolation, all data assimilation techniques used by Houser et al. [1998] are empirical. By empirical we mean that the estimation equations are neither derived from the optimization of an objective criterion, nor are they consistent simplifications such that the approximation error could be quantified in some way. An example for the former would be a full Kalman Filter, an example for the latter would be an eigenvalue decomposition with only the largest eigenvalues retained in the estimation algorithm. Optimal Interpolation in its full form could be called semi-empirical. It is an optimal interpolator in space at each isolated time step, provided the correct background error covariance is known. The suboptimality with respect to the Kalman Filter is due to the fact that the error covariance is not propagated. Therefore Optimal Interpolation is only optimal if the Kalman Filter happens to operate in steady-state and if the background error covariance of the Optimal Interpolation algorithm is equal to the steady-state Kalman Filter error covariance prediction. However, in the way Optimal Interpolation is implemented by Houser et al. [1998], namely through the ad hoc reduction of the number of measurements, the estimation algorithm is certainly empirical.

The degree of optimality in the data assimilation approaches implemented by Houser et al. [1998] is not clear and strongly depends on the particular application. In the case of nudging, the degree of optimality also depends on the choice of the many parameters. In the Optimal Interpolation scheme as implemented by Houser et al. [1998], the ad hoc reduction of the number of measurements defies a strict assessment of the approximation.

1.3.3 Soil Moisture Estimation from Atmospheric Observations

The Study by Mahfouf [1991]

Mahfouf [1991] introduced a technique to estimate soil moisture from the assimilation of low-level atmospheric parameters such as relative humidity and air temperature. The main purpose of the investigation is to come up with a better initialization of soil moisture in atmospheric models and consequently with better short- and medium-range weather forecasts.

The basic idea is that errors in the predicted meteorologic quantities may be related to errors in soil moisture. This is hypothesized to be particularly true for low-level air temperature and humidity, which are linked to surface and deep soil moisture by the sensible and latent heat fluxes. In other words, the assimilated data are not measurements of soil moisture, but rather observations of low-level atmospheric parameters, namely screen level temperature and relative humidity.

Two assimilation algorithms are developed. The first approach is a variational algorithm based on the minimization of an objective function. The objective function consists of the weighted sum of squared differences between the observed and the estimated low-level air temperature and relative humidity. No regularizing prior term for the uncertain initial soil moisture is included in the performance index, and the model is imposed as a strong constraint. The objective function is minimized with a standard Gauss-Newton method. This optimal scheme takes the modeled nonlinear relationship between the low-level parameters and soil moisture fully into account and is therefore computationally expensive.

The second approach is a statistical algorithm based on linear regression. This sequential scheme is suboptimal but computationally efficient and compatible with current operational assimilation. To get soil moisture estimates, the errors in the low-level atmospheric parameters are linearly related to the soil moisture errors in two layers, a very shallow top layer and a deeper reservoir. The linear relationship is described with a set of so-called nudging coefficients, which are in turn determined by an Optimal Interpolation analysis [Daley, 1991]. This implies that the nudging coefficients depend on the observation and forecast error statistics of the low-level atmospheric parameters as well as on the forecast error statistics of soil moisture. Mahfouf [1991] infers the necessary forecast error covariances with a Monte Carlo technique.

Mahfouf [1991] successfully assimilates field data into a one-dimensional version of a mesoscale numerical weather prediction model for three 48-hour periods. The results indicate that under certain atmospheric conditions it is indeed possible to estimate soil moisture from low-level atmospheric variables. The author also states that the variational scheme is preferable to the sequential scheme, but that the latter appears good enough to be given serious consideration for implementation in current operational numerical weather prediction systems.

The Study by Bouttier et al. [1993a]

Bouttier et al. [1993a] further investigate the sequential soil moisture estimation technique introduced by Mahfouf [1991]. As an alternative to the sequential approach of Mahfouf [1991], Bouttier et al. [1993a] provide an approximate analytic formulation of the nudging coefficients. This avoids the computationally expensive Monte Carlo simulations needed to come up with reasonable forecast error covariances. In the analytic formulation, the nudging coefficients are given as explicit expressions of the surface characteristics, most importantly of the vegetation parameters.

Moreover, Bouttier et al. [1993a] assess the sensitivity of the estimates to the vegetation, the soil texture, and the wind conditions. Vegetation is identified as the most crucial parameter. In a companion paper, Bouttier et al. [1993b] implement their simplified analytical filter in a mesoscale model. The study area covers $400 \times 400\, km^2$ in the southwest of France. Disturbing the initial soil moisture conditions from a reference simulation, Bouttier et al. [1993b] show that their nudging technique is able to restore soil moisture to the reference value within 48 hours.

The Study by Hu et al. [1999]

Hu et al. [1999] apply the sequential nudging technique of Mahfouf [1991] to 16 sites selected to sample a range of climates and land covers across the globe. Their goal is to derive a single set of nudging coefficients, which is then applied to a test site. The authors report computational instability when assimilating air temperature and relative humidity directly. Apparently, the strong correlation between air temperature and relative humidity makes the coefficient matrix of the equation for the nudging coefficients close to singular. The problem is overcome with a Principal Component Analysis.

Finally, Hu et al. [1999] conclude that numerical weather prediction can be improved by the nudging technique, but that nudging is unable to determine accurately the soil moisture within soil layers that are accessible to the atmosphere. Moreover, the improvement in the weather forecast holds only if the meteorologic model simulates precipitation poorly. If, on the other hand, precipitation is simulated well but surface radiation is modeled poorly, the nudging technique could erroneously adjust soil moisture.

The Studies by Callies et al. [1998] and by Bouyssel et al. [1999]

Following up on the variational approach of Mahfouf [1991], Callies et al. [1998] and Bouyssel et al. [1999] further investigate the feasibility of off-line soil moisture estimation for operational weather forecasting. The authors use one-dimensional versions of the soil and atmospheric boundary layer models of the German and the French weather services, respectively. Apart from minor differences, both soil models consist of a two-layer force-restore approximation for the soil temperature and of a two-layer soil moisture model. Callies et al. [1998] and Bouyssel et al. [1999] assimilate near-surface atmospheric measurements of air temperature and relative humidity to estimate the initial soil moisture conditions for a one-day and a two-day assimilation window, respectively. Whereas Bouyssel et al. [1999] choose ideal anticyclonic conditions with a clear sky and low advection, Callies et al. [1998] specifically choose a day reflecting non-perfect conditions.

While Callies et al. [1998] use the original strong constraint variational approach of Mahfouf [1991], Bouyssel et al. [1999] add a regularizing prior misfit term for the uncertain initial condition to the objective function. Neither study takes model errors into account. Note that the variational approach of Callies et al. [1998] is approximate because some of the fluxes in the atmospheric model are kept constant while the initial soil moisture is changed.

Both studies closely investigate the shape of the cost function and find that the initial soil moisture in the two layers cannot be estimated unambiguously. Lacking a regularizing term in the objective function, Callies et al. [1998] resort to fixing the soil moisture of the lower layer. Bouyssel et al. [1999] relate the ambiguity to the relatively short assimilation window and expect the problem to be solved by using longer assimilation periods.

In conclusion, the authors confirm the usefulness of the assimilation of soil moisture for short-term weather forecasts even under non-perfect conditions. Callies et al. [1998] state that *"a significant part of the information carried by the data cannot be explained by the need for higher energy input at the surface (stronger radiation) but must be attributed to an incorrectly modeled Bowen ratio probably resulting from a bad soil moisture specification"*. However, their retrieved soil moisture seems to be too low for the season, which reveals the estimated soil moisture as a tuning parameter for improving numerical weather prediction rather than a physical quantity.

The Studies by Rhodin et al. [1999]

In a recent study, Rhodin et al. [1999] apply the technique of Callies et al. [1998] to a regional weather forecast model. For the assimilation of soil moisture, all horizontal correlations are neglected and the three-dimensional problem is treated as a collection of completely entirely independent single-column assimilation problems. This offers huge computational savings, but large-scale structures in the errors of the soil moisture fields, arising for example from geologic processes, have to be neglected.

Discussion

Assimilating low-level atmospheric observations for the estimation of soil moisture offers great opportunities to improve short- and medium-range weather forecasts. Most importantly, the data are readily available within operational data assimilation system used for numerical weather prediction. However, the soil moisture values estimated in this way lack physical meaning. Indeed, Callies et al. [1998] deduce that *"soil moisture retrieval by the present method should be considered as a parametric approach"*.

Moreover, by dividing the domain into completely independent columns for the sake of soil moisture assimilation, the approach followed by Rhodin et al. [1999] does not allow for any explicit horizontal correlation of the initial condition of soil moisture. In their approach, soil moisture is correlated horizontally only through the spatial correlation of the low-level atmospheric parameters. This is clearly undesirable from a hydrologist's point of view and makes soil moisture even more of a tuning parameter.

Finally note that the indirect estimation of soil moisture from low-level atmospheric parameters is unsuitable for cloudy conditions or situations with predominantly large-scale advection. In these situations, air temperature and relative humidity are not related to local soil moisture.

1.4 Challenges in Hydrologic Data Assimilation

More research is necessary in order to set up an operational soil moisture data assimilation package. In particular, investigations should be undertaken along the following lines.

- The data assimilation approach should be truly optimal and four-dimensional, that is the large-scale correlation of land surface states should be considered explicitly. Introducing such structure only indirectly through correlations in low-level atmospheric parameters as done by Rhodin et al. [1999] is clearly inadequate from a hydrologist's point of view. Trying to estimate forecast state error covariances for Optimal Interpolation approaches is similarly unsatisfactory. In other words, the data assimilation algorithm should be optimal and provide for some form of large-scale error covariance propagation.

- In order to deal with the complexity of real world applications, a land-surface model suitable for hydrologic data assimilation has to be developed. Such a model must capture the key physical processes, but at the same time be highly computationally efficient.

- It is desirable to resolve the soil moisture profile in the field studies to a greater extent. A finer discretization in the vertical allows for a much better description of the nonlinear behavior than two-layer or three-layer models can provide.

- Off-line inversion of the remotely-sensed radiances into land surface states such as soil moisture can lead to inconsistencies in the model physics. It is therefore preferable to assimilate the remote sensing data directly into the hydrologic model.

- The temperature profile of the soil strongly affects the remotely sensed brightness temperature. When brightness temperatures are assimilated, it is necessary to model soil temperature along with soil moisture. The dynamics of the temperature profile can easily be described with the heat equation or approximations thereof, and observations of the surface temperature are readily available. Therefore, the land surface model should include soil temperature, and the estimation algorithm should provide for the assimilation of soil surface temperature measurements.

- Another problem that needs to be addressed is the mismatch between the scale of the hydrologic model and the scale of the observations. In particular, ground-based observations of soil moisture are generally point measurements, whereas remotely sensed observations are satellite footprints with a resolution of typically tens of kilometers. The same disparity in scales is true for precipitation measurements, one of the most important inputs for a soil moisture model. Moreover, inputs to hydrologic models are often available at scales finer than the scales of satellite remote sensing data. This creates a need for optimal downscaling methodologies. Hence, a consistent multiscale framework needs to be developed in order to accommodate measurements at different scales.

- Last but not least it is desirable to include more detailed models of the vegetation, as vegetation is probably the most important factor for the calculation of the latent and sensible heat fluxes at the land surface.

We address all of the above topics in our research. However, due to the highly nonlinear structure of the physical processes at the land surface, and given the high dimensionality of real world applications, a compromise will have to be made between a desirable physical foundation of the model and crude simplifications in order to achieve computational feasibility.